DOCUMENT RESUME

ED 225 822 SE 040 181

AUTHOR Herbert, Martin
TITLE Final Summary Report. Year 4. Evaluation Report 4-A-1. Extended Pilot Trial of the Comprehensive School Mathematics Program.
INSTITUTION CEMREL, Inc., St. Louis, Mo.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Dec 77
NOTE 37p.; For related documents, see SE 040 182-196 and SE 040 340-348.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Research; Elementary Education; *Elementary School Mathematics; Grade 3; Mathematics Achievement; *Mathematics Curriculum; Mathematics Instruction; *Program Effectiveness
IDENTIFIERS *Comprehensive School Mathematics Program; *Mathematics Education Research

ABSTRACT

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and was funded by the National Institute of Education (NIE). Its major purpose is the development of curriculum materials for the elementary grades. Beginning in 1973, CSMP ran an extended pilot trial of its elementary program. This report summarizes the data collected during its fourth year, with evaluation focusing on third-grade students and teachers. In the fall of 1976, CSMP materials were being used in 13 states and in about 1000 classes ranging from kindergarten to fourth grade. Evaluation here concentrates on the 23 districts and 108 classes using the third-grade materials. Three broad issues were explored: 1) After studying CSMP materials for 3 years, have students learned the computation skills and mathematics concepts usually expected of third graders? 2) Are these CSMP students better able to deal with certain novel mathematical problem situations than they would otherwise have been? and 3) What are the reactions of teachers to the CSMP third-grade materials? (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EXTENDED PILOT TRIALS OF THE COMPREHENSIVE SCHOOL MATHEMATICS PROGRAM:

EVALUATION REPORT SERIES

ED225822

Evaluation Report 4-A-1

Final Summary Report

Year 4

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
CEMREL INC.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE040181

Extended Pilot Trial of the Comprehensive School Mathematics Program

Evaluation Report 4-A-1

Final Summary Report

Martin Herbert

December, 1977

Developed by CEMREL, Inc., a private nonprofit corporation supported in part as an educational laboratory by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

Copyright on these materials is claimed only during the period of development, test, and evaluation, unless additional authorization is granted by the National Institute of Education, to claim copyright on the final materials. For information on the status of the copyright claim, contact either the copyright proprietor or the National Institute of Education.

Description of Evaluation Report Series

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and is funded by the National Institute of Education. Its major purpose is the development of curriculum materials for grades K-6.

Beginning in September, 1973, CSMP began an extended pilot trial of its Elementary Program. The pilot trial is longitudinal in nature; students who began using CSMP materials in kindergarten or first grade in 1973-74, were able to use them in first and second grades respectively in 1974-75, and so on in subsequent years. Hence the adjective "extended".

The evaluation of the program in this extended pilot trial is intended to be reasonably comprehensive and to supply information desired by a wide variety of audiences. For that reason the reports in this series are reasonably non-technical and do not attempt to widely explore some of the related research issues. The list of reports from the first two years of the extended pilot trial is given on the next page. The most comprehensive of these are the following:

1-A-1: Overview, Design and Instrumentation
1-A-3: Final Summary Report, Year 1
2-A-1: Final Summary Report, Year 2
3-B-1: Summary of Second and Third Grade Test Data Year 3
4-A-1: Final Summary Report, Year 4
and 4-A-2: Summary of Grades K-3 Evaluation

Evaluation Report Series

Evaluation Report 1-A-1	Overview, Design and Instrumentation
Evaluation Report 1-A-2	External Review of CSMP Materials
Evaluation Report 1-A-3	Final Summary Report Year 1
Evaluation Report 1-B-1	Mid-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-2	End-of-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-3	End-of-Year Test Data: Standard First Grade Content
Evaluation Report 1-B-4	End-of Year Test Data: CSMP Kindergarten Content
Evaluation Report 1-B-5	Test Data on Some General Cognitive Skills
Evaluation Report 1-B-6	Summary Test Data: Detroit Schools
Evaluation Report 1-C-1	Teacher Training Report
Evaluation Report 1-C-2	Observations of CSMP First Grade Classes
Evaluation Report 1-C-3	Mid-Year Data from Teacher Questionnaires
Evaluation Report 1-C-4	End-of-Year Data from Teacher Questionnaires
Evaluation Report 1-C-5	Interviews with CSMP Kindergarten Teachers
Evaluation Report 1-C-6	Analysis of Teacher Logs
Evaluation Report 2-A-1	Final Summary Report Year 2
Evaluation Report 2-B-1	Second Grade Test Data
Evaluation Report 2-B-2	Readministration of First Grade Test Items
Evaluation Report 2-B-3	Student Interviews
Evaluation Report 2-C-1	Teacher Questionnaire Data
Evaluation Report 2-C-2	Teacher Interviews, Second Grade
Evaluation Report 2-C-3	Teacher Interviews, First Grade
Evaluation Report 3-B-1	Second and Third Grade Test Data Year 3
Evaluation Report 3-C-1	Teacher Questionnaire Data Year 3
Evaluation Report 4-A-1	Final Summary Report Year 4
Evaluation Report 4-A-2	Summary of Grades K-3 Evaluation
Evaluation Report 4-B-1	Standardized Test Data, Third Grade
Evaluation Report 4-B-2	Mathematics Applied to Novel Situations (MANS) Test Data
Evaluation Report 4-B-3	Individually Administered Problems, Third Grade
Evaluation Report 4-C-1	Teacher Questionnaire Data, Third Grade

Key to Indexing

1-C-2 Observations of CSMP First Grade Classes

"2" refers simply to the number within a given year and type of data

"C" refers to the type of data being reported
- A: Overview, summary and theoretical reports
- B: Student outcomes
- C: Non-test data

"1" refers to the year of the Pilot Study according to the following:

	Kindergarten	First Grade	Second Grade	Third Grade	. . .
Year 1 (1973-74)	▨	▨			
Year 2 (1974-75)		▨	▨		
Year 3 (1975-76)			▨	▨	

Table of Contents

Introduction 1

Setting and Testing Plan 2

MANS Tests 5

Standardized Tests 10

Individually Administered Tests 13

Teacher Questionnaire 16

Summary 18

Appendix A Description of MANS Scales 21

Appendix B Graphs of Class Means on MANS Tests 25

Appendix C Individually Administered Problems: Section A 29

Introduction

In the fall of 1973, the Comprehensive School Mathematics Program (CSMP) began an Extended Pilot Test (EPT) of its Elementary School Program. Over one hundred teachers began using the program, either in first grade or kindergarten. During the 1974-75 and 1975-76 school years (the second and third years of the EPT) some of these classes continued with CSMP materials in the next grade levels and many new classes began using CSMP materials. The 1976-77 school year constituted the fourth year of the EPT, with some of the original classes continuing with CSMP in the third and fourth grades. Once again, many new classes began using CSMP at grade levels K through 4.

This report summarizes the data collected during this fourth year, in which the evaluation focussed on third grade students and teachers. Four separate evaluation reports[1] describe in more detail the data which will be summarized here; a larger summary of the whole set of evaluation studies[2] from grades K-3, covering 4 years of the CSMP Extended Pilot Test, relates this fourth year data to previous results.

[1] Evaluation Report 4-B-1 Standardized Test Data, Third Grade
" " 4-B-2 Mathematics Applied to Novel Situations (MANS) Test Data
" " 4-B-3 Individually Administered Problems, Third Grade
" " 4-C-1 Teacher Questionnaire Data, Third Grade

[2] Evaluation Report 4-A-2 Summary of Grades K-3 Evaluation

Setting and Testing Plan

In the fall of 1976, CSMP materials were being used in 13 states and in about a thousand classes ranging from kindergarten to fourth grade, 108 of them at the third grade level. Since the evaluation effort during this year of the Extended Pilot Test concentrated on the evaluation of the third grade materials, these 108 classes, in 23 school districts became the focal point of the evaluation.

Three broad issues were explored:

a) After studying CSMP materials for three years, have students learned the computation skills and the mathematics concepts usually expected of third graders?

b) Are these CSMP students better able to deal with certain novel mathematical problem situations than they would otherwise have been?

c) What are the reactions of teachers to the CSMP third grade materials?

It will be noted that the first two of these issues refer not to the effects of just the third grade materials, but rather to the cumulative effect of three years of using CSMP. In order to answer these questions the achievement of classes which had studied CSMP through third grade was compared with that of third grade classes which had not studied CSMP. In selecting places where such CSMP-Non-CSMP comparisons could be made, several considerations were important:

a) Classes studying CSMP materials for the first time in third grade were not used (for reasons given in the above paragraph).

b) Some classes dropped CSMP during the year.

c) In school districts with fewer than 4 third grade CSMP classes, it was decided that such small numbers of classes could not be representative of that school district. Furthermore statistical significance with such small numbers would require very large differences.

Altogether, 50 classes fell into one or more of these categories. This left 58 classes located in 8 school districts in which adequate comparisons might be made. All 8 initially agreed to participate in a comparative study; 3 of these 8 had to later withdraw for various reasons: One district decided to stop using CSMP materials the following year and hence had no further interest in evaluation data. A second district faced a financial squeeze which made school closing in early May a distinct possibility. The illness of a CSMP coordinator in a third district prevented another study from being carried out.

This left 5 school districts, containing a total of 33 third grade CSMP classes in which comparative studies were carried out. These 5 school districts are as shown below.

Table 1

Participating School Districts

School District	Site Number	Geographic Location	Type of Community	SES of Community	Number of CSMP Third Grade Classes
A	1	Midwest	Urban/ Suburban	Lower to Middle Class	4
B				Lower Middle to Middle Class	4
C				Upper Middle Class	4
D	2	Southeast	Rural/Small Town	Lower Middle Class	15
E	3	Northeast	Small City	Lower Middle to Middle	6

It can be seen that school districts A, B and C, which were suburbs of a large Midwestern City, were combined into one "site", insofar as these studies were concerned.

In Sites 1 and 3, the CSMP coordinators (district math supervisors) selected classes to serve as comparisons for the CSMP classes. These "Non-CSMP" classes were as comparable as possible to CSMP classes as far as ability level of class and SES of school were concerned. In Site 2 there were no comparable third grade Non-CSMP classes in the school district, but a very similar adjacent school district agreed to provide classes which would serve as comparison classes.

Three kinds of tests were administered in the spring to CSMP and Non-CSMP classes:

a) Mathematics Applied to Novel Situations (MANS) Test, a specially constructed series of 14 group administered test scales.
b) Standardized Mathematics Test
c) Individually Administered Problems

An ability test or reading test was also administered as a covariate at each site. The purpose of this test was to provide measures of ability for the various classes so that:

a) The comparability of CSMP and Non-CSMP classes could be inspected and statistical adjustments in class mean scores on the other tests could be made. As it happens, at each site the covariate scores for CSMP and Non-CSMP classes were quite comparable, and such adjustments were small.

b) The resulting F-test, upon which the statistical significance of CSMP-Non-CSMP differences is based, could be made more precise by the reduction of error variance made possible by the Analysis of Covariance Procedure.

Table 2

Testing Plan

Site Number	Number of Classes CSMP/Non-CSMP	Covariate Test	Standardized Mathematics Test	MANS	Individual Problems
1	12/12	Kuhlmann-Anderson	CTBS:[1] Computation	✓	✓
2	15/16	CTBS: Reading Comprehension	CTBS: Computation and Concepts & Applications	✓	Not Administered
3	6/12	ETS Cooperative Primary Reading	ETS Cooperative Primary Mathematics	✓	Not Administered

Essentially, there were seven separate studies:

a) The MANS data from each site
b) The standardized mathematics test data from each site
c) The Individual Problems from Site 1

The next three chapters describe the results from a), b), and c) respectively.

[1] Comprehensive Test of Basic Skills

Development of Scales

The MANS test is the 1977 version of a series of specially developed tests designed to gather data on the question: Are CSMP students better able to deal with certain kinds of novel mathematical situations than are Non-CSMP students? As in previous years, MANS/77 consists of a series of short scales, each one (a) involving a rather lengthy (½ to 6 min.) script read by the tester to orient the student to the novel situation, (b) consisting of an average of about 9 rather uniform items of increasing difficulty and (c) usually requiring no fixed time limit. The items themselves are usually open-ended, except on three scales involving estimation where a multiple choice format is used.

The criteria used for creating potential MANS scales were the following. (1) The task required is not covered in most standardized tests. (2) The language and/or content of the task-statement is not relatively more familiar to either CSMP or Non-CSMP students. (3) The task is recognized as a legitimate application of mathematics. (4) The task is thought to be related to a thinking process contained in the CSMP curriculum. Altogether 25 potential scales were drafted.

During the test development process each potential scale was reviewed by the CSMP Evaluation Panel (an external advisory group), the CSMP writing staff, and two school district supervisors who were coordinating CSMP in their district.

Through these reviews, and through extensive pilot tests, the scales were screened or revised in order to not only meet the four criteria stated above but also to pass procedural and empirical standards. In order for a scale to be retained, it had to be easily administered by a trained tester and the task easily understood by nearly all students. In addition, the scale had to be free of serious defects: e.g., uniformly easy or difficult items, evidence of much guessing, items in a single scale not testing the same ability, etc.

From the 25 scales which were originally developed, 14 were finally selected for use: scales A1-A7 in Booklet A and scales B1-B7 in Booklet B. Each of Booklets A and B were 16 pages in length and printed on newsprint and each required a total of 50-60 minutes administration time by a specially trained tester. Appendix A provides a brief description of these scales.

Two of the scales, A5 and B7, are quite different from the other scales. They dealt with straight computation skills not well covered in standardized tests at the third grade level. It was convenient to administer these scales as part of the MANS, but the scales certainly did not deal with novel applications and these results will be reported separately from the other MANS scales.

Results (excluding scales A5, B7)

The results are summarized in Table 3. The basic data are the adjusted class means; adjusted on the basis of the covariate score. Because the CSMP and Non-CSMP classes were always very close in ability, as measured by the covariate score, these adjustments at any site were never greater than 0.1 for any subtest.

The combined results in the right hand column of Table 3 show that CSMP classes scored significantly higher than Non-CSMP classes on both the "total" MANS/A and the "total" MANS/B. Looking at the 12 individual scales of the MANS type (excluding A5 and B7), it can be seen that CSMP classes scored higher than Non-CSMP classes in 11 of them, significantly so in 8 of the 11.

From the same Table 3, these across-site results can be checked on an individual site-by-site basis. On the "total" MANS/A and "total" MANS/B, the across-sites results are corroborated in Site 1 and Site 3. In Site 2, the results are still in favor of CSMP but not significantly so.

As a visual check for these results, graphs of class means were constructed for each scale, whereby each participating class was plotted; mean scale score versus mean covariate score. Inspection of these graphs showed that the linear regression model fit the data rather well; where there is a difference between CSMP and Non-CSMP classes it is consistent across class ability levels. Graphs for the total MANS scores are given in Appendix B.

Table 3
Adjusted Mean Scores, MANS Scales

First Entry: CSMP Classes, Second Entry: Non-CSMP Classes

*, **: Significant at .05, .01 levels respectively [2]

MANS A	Site							
	One		Two		Three		Combined [3]	
Number of Classes	12	12	7	6	6	12	25	30
Raw Score, Covariate [1]	69.5	68.6	28.0	28.5	35.6	36.7	54.5	55.9
A1. Height and Weight Table	3.5	3.4	2.9	2.9	3.3	3.7	3.3	3.4
A2. Estimation	10.7	10.3	8.7	8.7	11.7	9.9	10.5	9.7 *
A3. Functions	4.4	3.4 **	2.9	2.5	4.9	3.8 **	4.2	3.4 **
A4. Two Stage Word Problems	3.1	2.6 **	2.7	2.2	3.2	3.1	3.0	2.7 **
A5. Large Number Computations	6.1	5.3 **	4.6	4.4	6.4	5.3 **	5.9	5.0 **
A7. Hints and Problems	2.1	1.8	1.3	1.3	2.2	1.7 *	2.0	1.6 **
Total	28.5	26.2 **	22.5	21.7	30.9	27.1 **	27.8	25.3 **
MANS B								
Number of Classes	12	12	8	6	6	11	26	29
Raw Score, Covariate [1]	69.5	68.6	28.0	28.0	35.6	36.7	54.5	56.7
B1. 2 or 5 or 10	6.3	5.8 *	5.3	5.2	6.4	6.1	6.1	5.7 *
B2. Composite Functions	6.4	5.5 **	5.0	4.4	7.3	5.7 **	6.3	5.3 **
B3. Equation Fluency	6.0	6.1	5.2	4.8	6.5	6.2	5.9	5.8
B4. Circle the Larger	9.3	7.9 **	8.0	7.5	9.3	8.2 **	8.9	7.9 **
B5. Missing Digits	3.7	3.0 **	2.9	2.9	3.0	3.1	3.3	3.0 **
B6. Word Problems "with Rounding"	2.3	2.2	1.9	1.6	2.5	2.4	2.3	2.1
Total	34.0	30.5 **	28.2	26.4	35.0	31.7 **	32.8	29.8 **

[1] Kuhlmann Anderson, CTBS Reading Comprehension and ETS Reading for Sites 1,2,3 respectively. Transformed percentile ranks for Combined.

[2] Significance tests were based on a one-way Analysis of Covariance procedure, with 1 and 21, 1 and 10, 1 and 15, and 1 and 52 degrees of freedom for Sites 1,2,3 and combined respectively.

[3] Because a different covariate score was used at each site, the raw scores were transformed into percentile ranks that were comparable from site to site; i.e., site-to-site differences were reduced as much as possible in order to concentrate on CSMP-Non-CSMP effects.

One can legitimately ask if the results obtained using class means as the unit of analysis will hold when data is aggregated across individual students. For example is it also true if one looks at low ability students, that roughly the same results will appear as with high ability students? To answer this question the following procedure was used:

Based on his or her covariate score, each student was assigned to either the first quintile (for students below the 20th percentile rank), or the second (20th-40th), or third (40th-60th), or fourth (60th-80th) or fifth quintile (above 80th percentile rank). Then, in each quintile, the average MANS score was computed for CSMP students and for Non-CSMP students. These mean scores are shown in numerical and graphical form in Figure 1 and 2.

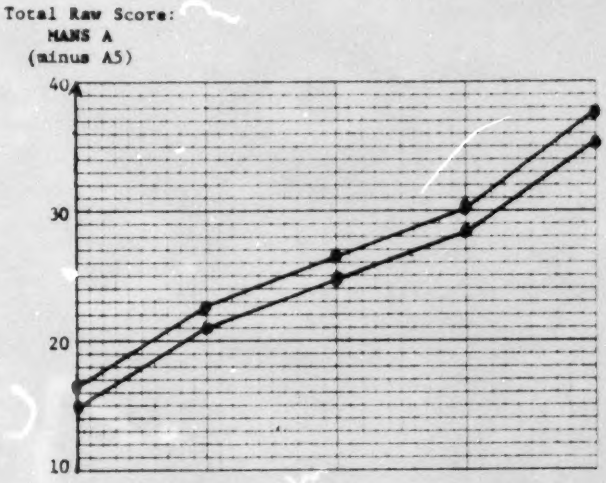


Figure 1

MANS/A Performance by Covariate Quintiles

Range of Percentile Ranks:	under 20	20-40	40-60	60-80	over 80
Mean, CSMP Students:	16.4	22.8	26.5	30.2	37.8
Mean, Non-CSMP Students:	14.9	21.0	24.6	28.2	35.4

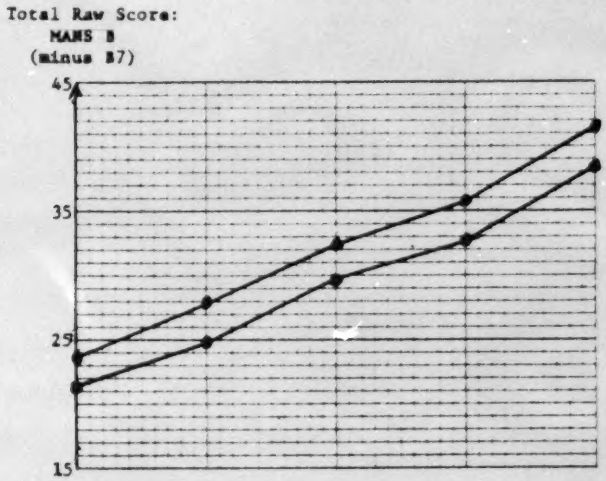


Figure 2

MANS/B Performance by Covariate Quintiles

Range of Percentile Ranks:	under 20	20-40	40-60	60-80	over 80
Mean, CSMP Students:	23.7	27.9	32.5	35.8	41.3
Mean, Non-CSMP Students:	21.1	24.8	29.9	32.7	38.3

As can readily be seen, the CSMP advantage is consistent across each of the five ability levels for both the MANS tests. When the same information was developed for each individual scale, it was found that this pattern holds true for each of the eight scales in which there is a significant difference. In addition, in the three scales where the CSMP score is higher but not significantly so, in no quintile do Non-CSMP students outscore their CSMP counter-parts. Clearly it is the case that the general results hold true for all ability levels.

Results for Scales A5 and B7

Scale A5 dealt with computations involving larger numbers (such as 3 x 125 = ☐) and Scale A7 dealt with fractions, specifically $\frac{1}{2}$ or $\frac{1}{3}$ of a whole number. These are straight computation scales containing items not found in standardized tests at the third grade level. It was expected that CSMP students would do better on these scales; not because of any superiority in "applying math to novel situations" but because they had more experience in these kinds of computations than did Non-CSMP students.

The results for these two scales are given below in Table 4. As expected, the CSMP classes did much better than Non-CSMP classes, significantly so on all comparisons with Scale B7 and in all comparisons except Site 2 with Scale A5.

Table 4

Adjusted Class Means: Scales A5, B7

First Entry = CSMP Classes; Second Entry = Non-CSMP Classes

	Site 1		Site 2		Site 3		Combined	
A5 = Large Number Computation	6.1	5.3	4.6	4.4	6.4	5.3	5.9	5.0
B7 = Fractions	4.1	2.8	2.9	2.0	4.8	3.3	4.0	2.8

Standardized Tests

Although the objectives of CSMP have never included, as a major concern, the improvement of student performance in what might be called the "usual" computational skills and mathematics concepts, it is true nevertheless that the program expects that its students will do at least as well in these areas as they would have done in a more traditional mathematics program. Throughout the Extended Pilot Trial, standardized tests have been used to compare CSMP and Non-CSMP students' performance on these traditional skills and concepts.

To review, the following standardized math tests were administered:

At Site 1, The Computation Test of the Comprehensive Tests of Basic Skills (CTBS)
At Site 2, a) The Computation Test of the CTBS
b) The Concepts and Application Test of the CTBS
At Site 3, The Mathematics Tests of the Cooperative Primary Test (ETS)

A separate Analysis of Covariance procedure was carried out at each site. Table 5 shows, for each test, adjusted mean scores across CSMP and Non-CSMP classes. Asterisks denote significance at the .05 level. The main findings are as follows:

1. There was a significant difference in favor of Non-CSMP classes at Site 2 on the CTBS Concepts and Applications Test.
2. There were no significant differences between CSMP and Non-CSMP classes:
   a) at Site 1 on the CTBS Computation Test
   b) at Site 2 on the CTBS Computation Test
   c) at Site 3 on the ETS Cooperative Mathematics Test
3. On the subsets of items within the CTBS Computation test, at both Sites 1 and 2 CSMP students did better on addition and multiplication items and Non-CSMP students did better on the division items. These differences were significant at Site 1.

Each of the above results held true at each level of student ability as well as across classes generally.

Table 5

Adjusted Mean Scores, Standardized Math Tests

(*, **: Significant at .05, .01 levels respectively)[2]

	Site 1		Site 2		Site 3	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
Comprehensive Tests of Basic Skills:						
Number of Classes	12	12	15	13		
Raw Score, Covariate[1]	69.7	68.7	27.9	27.8		
Addition items	9.9 **	9.2	9.7	9.6		
Subtraction items (12)	9.4	8.7	8.6	9.2		
Multiplication items (12)	9.6 **	8.8	9.0	8.8		
Division items (12)	7.3	8.7 **	8.4	9.0		
Total Computation (12)	36.1	35.4	35.8	36.8		
Concepts (25)			16.6	18.0 **		
Applications (25)			14.7	15.8 *		
Total Concepts & Applications			31.2	33.8 **		
ETS Cooperative Primary Math						
Number of Classes					6	12
Raw Score, Covariate[1]					35.7	36.7
ETS Math					42.3	41.8

[1] Kuhlmann Anderson, CTBS Reading Comprehension and ETS Reading for Sites 1, 2 and 3 respectively.

[2] Degrees of freedom for F-test were 1 and 21, 1 and 24, and 1 and 15 for Site 1, 2 and 3 respectively.

The finding that Non-CSMP classes did better on the Concepts and Applications Test of the CTBS at Site 2 is inconsistent with other data (from both this year and previous years) in which CSMP classes have generally done as well as or better than Non-CSMP students. Because of this inconsistency, the Site 2 results were investigated in some detail. In particular, for each item, the percentages of CSMP and Non-CSMP students choosing the correct answer were calculated. Similar information was also available from Site 1 third grade classes last year and from a few of the 24 Site 1 classes this year (not reported elsewhere because of the non-comparability of CSMP and Non-CSMP classes). Thus it was possible to compare, item-by-item, CSMP-Non-CSMP differences at Site 2 versus CSMP-Non-CSMP differences at Site 1.

The main findings on the 50 items of the test are as follows:

1) At both sites, CSMP classes did clearly better than Non-CSMP classes on three unrelated items
2) At both sites, CSMP classes did clearly worse than Non-CSMP classes on five items, three of which dealt with measures of one kind or another: clocks, calendars, and feet/inches.
3) There were 10 items on which CSMP students at Site 1 did much better but CSMP students at Site 2 did much worse than Non-CSMP students. Six of these 10 dealt with either set pictures or number sentences (equations).

Other than the findings in favor of Non-CSMP at Site 2 on the Concepts and Applications Test, the results were similar to those obtained in previous years: little or no difference between CSMP and Non-CSMP classes at the total test level, better performance by CSMP classes on addition and multiplication items and inconsistent results on subtraction items. The items on division were done better by CSMP students last year but better by Non-CSMP students this year.

Individually Administered Problems

In order to get a better understanding of possible differences between CSMP and Non-CSMP students in the way they went about various mathematical processes, two individually administered sets of problems, A and B, were developed. The study was exploratory in nature. Nevertheless the process of selecting students and administering and scoring these problems was standardized to a certain extent, allowing gross comparisons between CSMP and Non-CSMP students to be made. Some numerical data will be reported in this chapter, but this should be viewed as indicative of possible curriculum differences. Significance tests were not performed. The actual problems are outlined very briefly below; further details can be found in Appendix C.

The problem sets were designed to be used with pairs of students, the students alternating in answering questions and occasionally cooperating (or competing). Each problem set usually required about 45 minutes per student pair. A total of 35 representative pairs of students were selected from 10 CSMP classes in Site 1. Usually 5 or 6 pairs of students were selected from each school where CSMP was being used (1 or 2 classes). Half the pairs went through Section A, the other half, Section B. These pairs were matched, school-by-school with pairs of Non-CSMP students who had equivalent covariate scores.

Section A: Numerical Processes

For Part I each student was given a set of 18 problems which had been "completed" i.e., the answer box had been filled in by "someone". Each student went quickly through them and beside each problem put either an "x", for pretty sure it's wrong, or a check, for could be right. After this was done each problem was reviewed and if an "x" had been recorded, the student was asked to explain why he or she was pretty sure the answer was wrong (13 of the 18 answers actually were wrong). Each explanation was classified according to whether or not the student gave an adequate reason for the given answer being wrong or a good estimate of what the answer should have been (versus an incorrect or insufficient explanation). Special follow-up questions were designed for certain problems.

Thus Part I contained three tasks: initial determination of correctness of "completed" answer, explanation of why completed answer was wrong, and follow-up questions specific to a given problem.

Part II consisted of 7 problems which the students worked on together, orally, with the interviewer. The problems dealt with estimation; pattern recognition; relative sizes of large numbers; and relating the processes of addition, multiplication and division. Hand calculators were used in two of the problems. A scoring protocol was developed for Parts I and II and the mean scores obtained are summarized in Table 6, below. Assignment of students to ability level in Table 6 is based on covariate score.

Table 6

Mean Scores, Section A

	Ability Level of Students							
	High		Medium		Low		Combined	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
Part I	24.1	23.9	21.5	16.4	14.1	14.6	20.3	17.4
Part II	15.7	14.8	15.7	12.4	12.4	9.3	14.7	12.1

It can be seen that for Part I, CSMP students at the middle ability level had higher scores than their Non-CSMP counter parts. For Part II, CSMP students at both the middle and low ability levels scored higher.

Section B: Concept Identification and Classification

There were 5 parts to Section B. Each part will be described briefly below, accompanied by a summary statement of the main finding.

Part I: The task was to find the Interviewer's "secret number", between 1 and 100, by asking as few YES-or-NO type of questions as possible, i.e. by being as efficient as possible in eliminating possibilities. Each pair did this four times, twice for each student. There were several ways that the data were analyzed (number of questions required, type of elimination strategy used, point at which single-number guessing began), but all lead to essentially the same conclusion: CSMP students were better at this problem on their first trial, but Non-CSMP students improved to a comparable level of performance on the second.

Part II: For this and subsequent parts of Section B a 16 piece set of 4 x 6 inch "People Pieces" was used. Each piece showed either a boy or girl, who was either tall or short, fat or thin, and red or blue. Students were told to put the pieces into piles "so that all of the cards in one pile are the same in some way and so that each pile is different from the others". Students took turns until neither could think of any other way to classify the cards. There was no difference between CSMP and Non-CSMP students in the number of sorts made, although CSMP students on the whole tended to sort the pieces into more complex categories (i.e., simultaneously involving more dimensions of sex, color, height and width).

Part III: The interviewer had a "secret rule" involving one or more dimensions of the "people pieces". The student tried to figure out what the rule was by offering pieces one after another, each of which was placed in either a YES column (if the piece fit the rule) or a NO column (if it didn't). Each of the pair of students had 2 trials; in one the secret rule had a single dimension (such as "short") and in the other it had two dimensions (such as "fat and tall"). There was no difference between CSMP and Non-CSMP students in the average number of pieces needed to solve the single dimension rules, but CSMP students required only about 60% as many pieces as did Non-CSMP students for the two dimension rules.

Part 4: This was similar to Part 3 except that instead of the student deciding which pieces to offer, the student was shown a standardized sequence of pieces "someone else had tried" and which column (YES or NO) each had been placed in. After three pieces, students were asked what the rule could possibly be. There was no difference between CSMP and Non-CSMP students in the number of suggestions made (about 3) or in the number of these which contradicted the given information.

Part 5: This was also similar to Part 3, except that a third color was introduced by adding a set of green pieces to the People Pieces, and instead of a YES and a NO pile, there was merely a NO pile. Pieces fitting the secret rule were merely turned face down, so that the student could see in front of him only a set of pieces which did not fit the secret rule. About 83% of both the CSMP and Non-CSMP students eventually figured out the rule, although Non-CSMP students required slightly fewer pieces (an average of 5.4 versus 6.0).

Teacher Questionnaire

In the spring, 1977, questionnaires were sent to 69 third grade CSMP teachers and were completed and returned by 59% of them. Over half these teachers were also interviewed. Some of them had taught CSMP before and were working with students who had studied CSMP before. Other teachers were using CSMP for the first time; sometimes with students new to CSMP, at other times with students who had studied CSMP before. Teachers who returned the questionnaire appeared to be fairly representative of the total group of third grade teachers in the Extended Pilot Trial in terms of experience and overall attitude towards CSMP.

The responses will be summarized according to three areas: Overall Judgments, Student Achievement and Mechanical Aspects.

Overall Judgments

On an open-ended question asking for an overall evaluation of CSMP, 86% of the responses were generally positive towards CSMP, although 2/3 of these expressed reservations of one kind or another.

On an open-ended question on the spiral approach of CSMP, about twice as many responses were positive as were negative (about a third of the responses were mixed).

In comparing CSMP to the previously used mathematics curricula, three times as many responses indicated a higher, as opposed to a lower, level of student enthusiasm and three times as many responses indicated that the overall quality of CSMP was higher rather than lower (about a fifth of the responses were neutral).

Student Achievement

There are many indications that teachers see a need for more emphasis on basic skills:

In comparing CSMP with previous math curricula, teachers were evenly divided as to whether they thought CSMP was better or worse as far as student achievement of the usual skills and concepts was concerned. The same thing was true of students' facility in solving word problems.

About 3/4 of the teachers supplemented the program on a regular basis, usually with practice in the basic computational skills.

Over half the teachers noted that there were some basic skills that their CSMP students should have learned but had not.

The most frequently mentioned "worst aspect" of CSMP was that there was not enough drill on the basic facts.

Closely related to the basic skills issue is the appropriateness of CSMP for the low ability students. Almost 2/3 of the teachers felt it was less appropriate than their previously used math program, and this was the second most listed "worst aspect" of CSMP. More negative responses came from teachers of high, as opposed to low, ability classes.

On the other hand, many teachers noted that CSMP helped children to think and be creative; and most teachers thought CSMP students' ability to do logical reasoning was better than with their previous math program.

Mechanical Aspects

Teachers generally seemed to be able to cope with the program; most followed the schedule and expected to complete it by the end of the year. There were many fewer references to difficulties in managing materials and not knowing what to do than in previous years and grade levels, though most teachers again expressed a need for built-in tests in the materials.

Regarding time requirements, most teachers felt that (1) CSMP would not require more preparation time than usual after a year's experience, (2) more class time was needed for math instruction and that this was a positive thing, and (3) more of the class time was teacher-led and that this also was a positive thing. A sizable minority (about 25%) disagreed with (1) and thought (2) and (3) were unfavorable aspects of CSMP.

Most teachers felt reasonably well prepared for CSMP after their training (again a sizable minority disagreed) and thought that most third grade teachers could teach it adequately. Meeting informally with other teachers seemed to happen more frequently, and be more useful, than meeting with the coordinator during the course of the year.

Summary

Questionnaire Data: Teachers were positive in their overall attitudes towards CSMP, especially regarding student enthusiasm, student ability to do logical thinking, and overall quality of CSMP. However, more than half the teachers expressed concern about the appropriateness of the program for low ability students and about the lack of sufficient drill on the basic facts. Compared to previous questionnaire data, these teachers expressed less favorable attitudes towards the program as a whole, but seemed to have fewer problems in teaching it.

Individually Administered Problems: Each of two problem sections was administered to about 18 representative CSMP and Non-CSMP pairs of students. The study was exploratory in nature and significance tests were not performed, but there were indications that, at least on a few of the problems, there were real differences in favor of CSMP students.

Group Administered Tests: Table 7 summarizes the test data collected from i) the MANS Tests, a series of short test scales containing mathematical tasks which are novel to third graders, but thought to be related to some of the thinking process of the CSMP curriculum, and ii) Standardized Tests.

Table 7
Summary of Adjusted Mean Scores
(* = significant at .01 level)

Test	Site 1		Site 2		Site 3	
	CSMP Classes	Non-CSMP Classes	CSMP Classes	Non-CSMP Classes	CSMP Classes	Non-CSMP Classes
Number of Classes	12	12	15	13	6	12
Mean Raw Score, Covariate Test	69.7	68.7	27.9	27.8	35.7	36.7
MANS A (excluding A5)	28.5*	26.2	22.5	21.7	20.9*	27.1
MANS B (excluding B7)	34.0*	30.5	28.2	26.4	35.0*	31.7
Special MANS Computation Scales (A5 + B7)	10.2*	8.1	7.5*	6.4	11.2*	8.6
CTBS Computation	36.1	35.4	35.8	36.8		
CTBS Concepts and Applications			31.2	33.8*		
ETS Cooperative Primary Mathematics					42.3	41.0

a) MANS Tests

CSMP classes did better on the MANS Scales, including the two special computation scales (Fractions and Large Number Computation). These scales were developed with CSMP in mind so that a result favoring Non-CSMP classes would have been very unexpected. However, with the exception of the computation scales (which were content-biased in favor of CSMP), the tests should not have favored CSMP students merely because they were more used to doing these specific tasks. Neither the language and terminology of CSMP, nor the problem exercises of the CSMP curriculum, appeared in any of those scales. If they were biased in favor of CSMP it was because the CSMP curriculum had developed in these students a better facility for dealing with problems of this kind; this would be bias in the best sense of the word. The results obtained this year on tests of this kind are similar to, but stronger than, results from previous years. In fact there seems to be a discernable trend of gradually increasing differences on tests like these in favor of CSMP in higher grade levels (i.e., with more exposure to CSMP).

Regarding the MANS results, it is worthwhile exploring in some detail the validity of the results. Was it the CSMP curriculum which caused differences in achievement in favor of CSMP students or is some other explanation just as tenable? The three most plausible rival explanations appear to be the following:

1) CSMP classes spent more time on math than Non-CSMP classes,
2) There was a Hawthorne effect,
2) CSMP teachers were systematically different from, and in sense better than, Non-CSMP teachers.

Items 1) and 2) have been investigated this year and/or in previous years. Regarding 1) it was possible to compare the amount of class time CSMP and Non-CSMP teachers reportedly spent in math instruction, and there was no difference. Previous studies have shown little or no differences. Regarding 2), a previous study has shown that CSMP teachers in fact do a better job the second year than the first, i.e,, added experience more than compensates for any loss in enthusiasm. (Of course it may be that enthusiasm is sustained for 2 or 3 years before waning, so one cannot entirely rule this out.)

As regards item 3), which is probably the most plausible, the following argument applies. There were altogether 13 schools which contained the 33 participating CSMP classes and in 11 of these schools it was the case that every third grader in the school studied CSMP math. This means that within schools there was no choice of assigning students or teachers to CSMP or not to CSMP; they all used it as the regular math program. This reduces considerably the possibility that CSMP teachers and students were systematically better tnan Non-CSMP teachers becasue of volunteerism or deliberate selection. In effect the unit of selection was the school and one has to ask whether the CSMP schools had systematically better third grade teachers and students than did Non-CSMP schools. This seems unlikely because: a) the schools were usually nearby and in neighborhoods of approximately the same SES level, b) the distribution of clsss mean covariate scores was very similar for CSMP and Non-CSMP at each site, and c) the selection to use CSMP by the coordinator of a school would, if anything, depend on the enthusiasm of first grade teachers (to get the program "started" right); the third grade teachers more or less inherit both the students and the CSMP curriculum.

b) Standarized Tests

The only other entry in Table 7 which is significant, is on the Concpets and Applications section of the CTBS, where Non-CSMP classes did significantly better, and this result is not similar to results obtained in previous years. In fact the most ususal result on standardized tests has been "no significant differences", and this is what happened at Sites 1 and 3, and on the CTBS Computation Test at Site 2. Whether the aberrant finding is a warning of a potential trouble spot, or merely an example of normal site-to-site differences which should be expected, remains to be seen. At this time one tends to discount it because of the mass of evidence from other years, other sites and other tests, which indicates that students in CSMP do quite well, at least through the first few years of the program.

Generalizability of Results: Should one expect the various results from this year's Extended Pilot Test tc occur in third grades of other school districts? The participating school districts seem to be representative of many school districts, but ovbiously they were not random samples from the target population of (all) school districts. They were merely school districts which were a) among the larger users of CSMP at third grade and b) interested in collecting evaluation data. Thus the answer to the above question must be "not necessarily", but the results are nevertheless rather promising.

Appendix A

Description of MANS Scales

For each scale the following information is given: (1) the scale letter/number, its name, and the number of items, (2) an abstract of the content, and (3) a relatively easy sample item. There is an asterisk placed before the two scales not of the MANS-type.

SCALE A1: Height and Weight Table (6 items)

ABSTRACT: Read and interpret data from a table of students' weights and heights for two different years

SAMPLE: Who stayed the same height? (Note to the reader: the table can be seen in Appendix B)

SCALE A2: Estimation (25 items)

ABSTRACT: Quickly estimate which of 5 standard intervals contains the answer to each of a series of computation problems. Three separate pages containing 8 addition, 8 subtraction and 7 multiplication problems respectively.

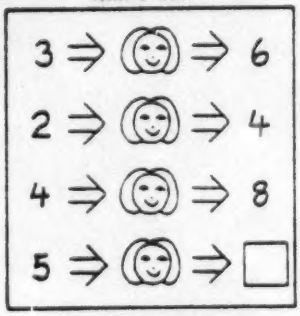
SAMPLE: 100 - 93 0 10 50 100 500 1000

SCALE A3: Functions (8 items)

ABSTRACT: For each of several problems, determine from 3 pairs of numbers what the "secret rule" is which produces the second number from the first, and use it to find the missing number from the 4th pair.

SAMPLE:

Kim's Game



SCALE A4: Two Stage Word Problems (5 items)

ABSTRACT: Word problems (printed in booklet and read by tester) in which two different operations must be performed and where the numbers in the given data are relatively small.

SAMPLE:

Our hens lay 9 eggs every day.

Each day we eat 6 of them and give the others away.

During the next 5 days how many eggs will we give away? _____

* SCALE A5: Large Number Computations (12 items)

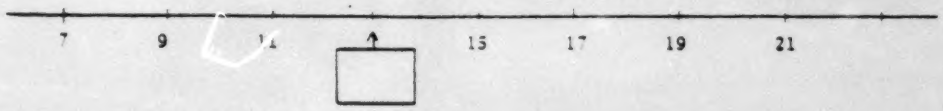
ABSTRACT: Solve computation problems given in an open sentence format, with the boxes sometimes in non-standard positions, and with numbers in the hundreds but relatively easy to work with (addition, subtraction and multiplication).

SAMPLE: $500 + \square = 800$

SCALE A6: Number Line Labelling (8 items)

ABSTRACT: Label the indicated "mark" on several number lines, where marked intervals vary from item to item.

SAMPLE:



SCALE A7: Hints and Problems (5 items)

ABSTRACT: Quickly complete a given addition problem by using the answer to another problem where one addend is the same as, and one is only slightly different from the given problem.

SAMPLE: Hint: $537 + 293 = 830$

$537 + 283 = \square$

SCALE B1: 2 or 5 or 10 (10 items)

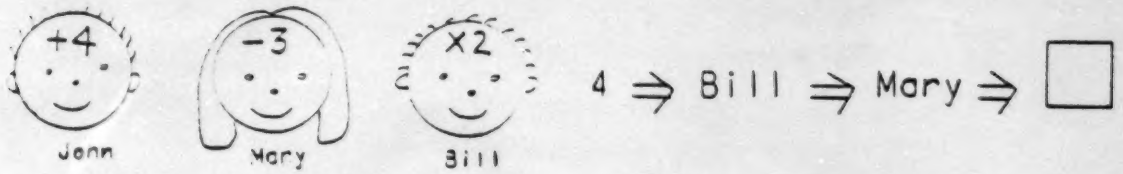
ABSTRACT: Quickly estimate whether a given number is about 2 or 5 or 10 times as big as another given number.

SAMPLE: 60 is about ____ times as large as 31

SCALE B2: Composite Functions (9 items)

ABSTRACT: Starting with a given number, apply one or more operations in sequence and determine final result. Also, same process except final result is known and starting number is to be determined.

SAMPLE:



SCALE B3: Equation Fluency

ABSTRACT: Given the symbols: = + - x 1 2 3 ( ), construct as many different equations as possible

SAMPLE ANSWER: 3 - 1 = 2

SCALE B4: Circle the Larger (13 items)

ABSTRACT: Given pairs of computation problems, quickly determine which one has the larger answer.

SAMPLE: 371 + 248 370 + 258

SCALE B5: Missing Digits (8 items)

ABSTRACT: Given a computation problem with one or two digits of the problem crossed out, determine whether or not the given answer could have been right (before the digits were crossed out).

SAMPLE:

$$\begin{array}{r} 54 \\ +3\blacksquare\blacksquare \\ \hline 500 \end{array}$$

Could 500 be the answer?

No, 500 is too small. ☐
Yes, 500 could be right. ☐
No, 500 is too big. ☐

SCALE B6: Word Problems with "Rounding" (5 times)

ABSTRACT: Solve word problems (printed in booklet and read by the tester) involving division in which the given numbers do not divide evenly - i.e., the answer, which must be an integer, can be obtained by rounding, up or down. The numbers of the given data were relatively small.

SAMPLE:

An elevator can't hold more than 5 people.

23 people want to ride to the top floor.

How many times will the elevator have to go up? _____

SCALE B7: Fractions (8 items)

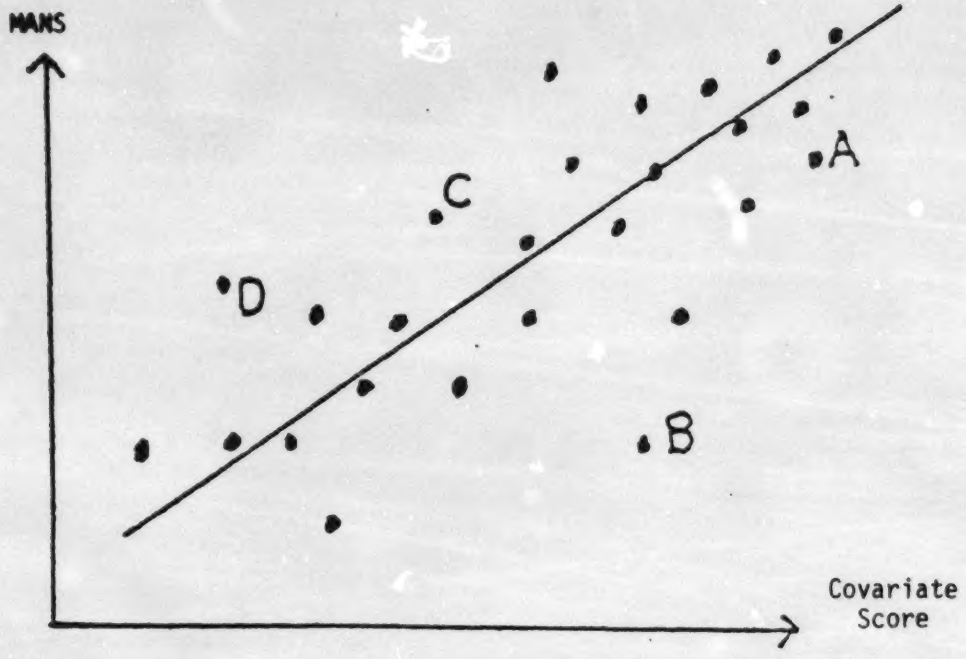
ABSTRACT: Solve problems of the form x of y = ☐ or x of ☐ = y where x is 1/2 or 1/3.

SAMPLE: $\frac{1}{2}$ of 12 = ☐

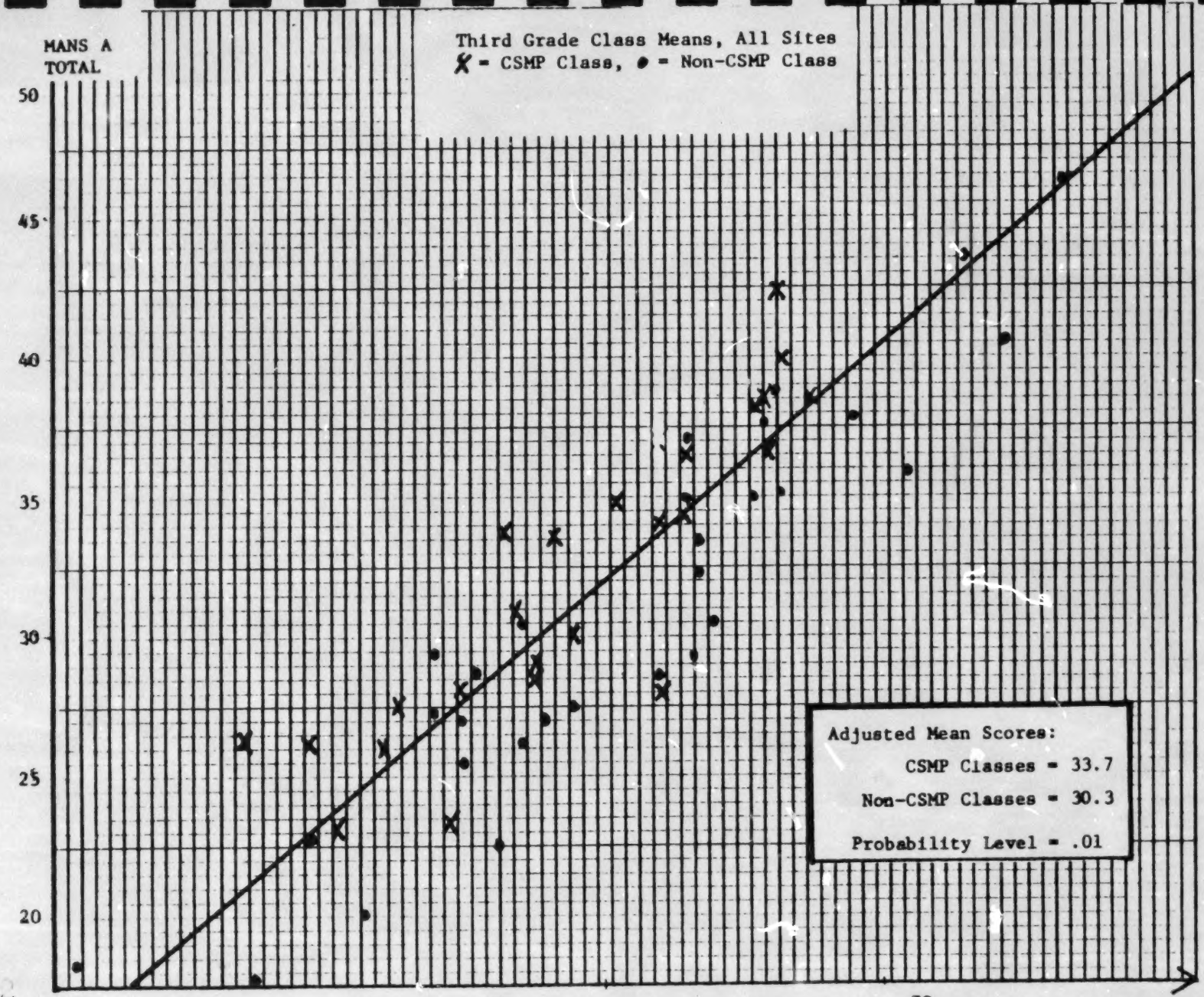
Appendix B

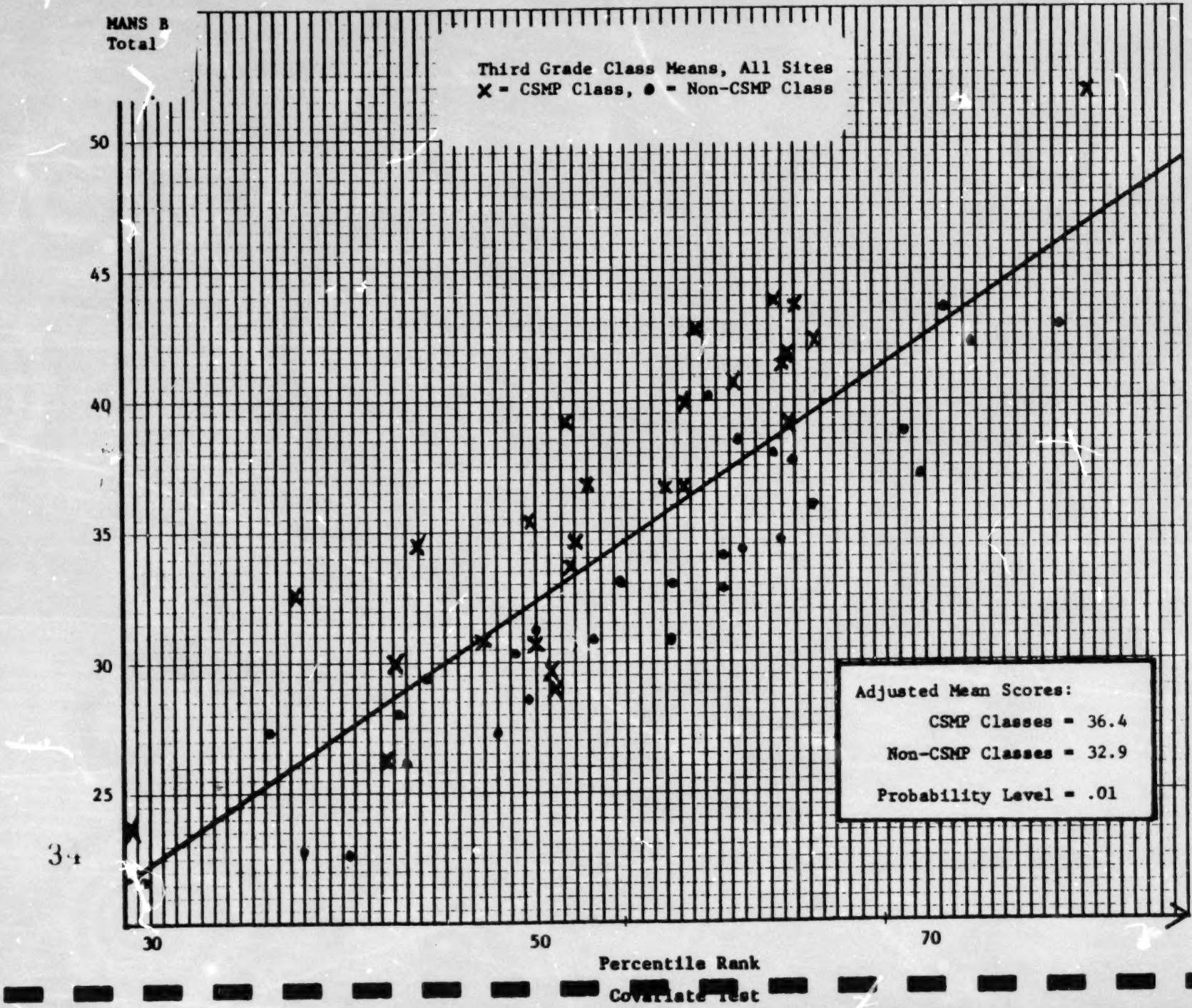
Graphs of Class Means on MANS Tests

The graphs of class means are in the form shown below.



The vertical axis is for Total MANS score; the further towards the top of the graph, the higher the class score. The horizontal axis is for covariate score; the further to the right, the higher the general ability of the class. Based on the set of class means thus graphed, the regression line has been drawn. This line is the best linear prediction of mean class test score that can be made from knowing the ability level of the class. (Note in the example that classes A and B fall well below the regression line, or are scoring well below what would be predicted for them from knowing the ability level of the class, while classes C and D fall well above the regression line. Note also that, although class A had a slightly higher mean score on the test than did class C, class C did much better given the relative ability scores of the two classes.) Throughout, an x is used to represent a CSMP class and a dot (.) represents a Non-CSMP class. In an unused corner of each graph, a table gives the adjusted Mans scores across CSMP classes and across Non-CSMP classes, together with the p-value obtained from the F-test.


MANS A
TOTAL
Third Grade Class Means, All Sites
X = CSMP Class, • = Non-CSMP Class
50
45
40
35
30
25
20
Adjusted Mean Scores:
CSMP Classes = 33.7
Non-CSMP Classes = 30.3
Probability Level = .01
30
50
70
Percentile Rank
Covariate Test


MANS B
Total
Third Grade Class Means, All Sites
X = CSMP Class, ● = Non-CSMP Class
50
45
40
35
30
25
30
50
70
Percentile Rank
Covariate Test
Adjusted Mean Scores:
CSMP Classes = 36.4
Non-CSMP Classes = 32.9
Probability Level = .01

Appendix C

Individually Administered Problems: Section A

Part I: Set of problems which had been "completed" by "someone"
(Student copies had handwritten, rather than typed, answers within boxes)

$26 + 36 = \boxed{512}$

$6 \times 13 = \boxed{53}$

$\boxed{47} - 14 = 33$

$58 + 62 = \boxed{120}$

$30 + 40 + 50 = \boxed{125}$

$\boxed{192} + \boxed{192} + \boxed{192} + \boxed{192} = 428$

$\frac{1}{2}$ of $316 = \boxed{158}$

$24 + \boxed{47} = 91$

$3{,}186 + 5{,}279 = \boxed{9{,}315}$

$6{,}473 + 3{,}412 = \boxed{985}$

$4 \times 106 = \boxed{824}$

$\boxed{30} - 43 = 13$

$\boxed{98} + \boxed{98} + \boxed{98} = 294$

$8 \times \boxed{12} = 160$

$\frac{1}{2} \times 92 = \boxed{56}$

$\boxed{158} + \boxed{158} = 316$

$\boxed{600} + 3 + 13 + 7 + 8 = 731$

$53 + \boxed{20} = 33$

Part II: Summary of Problems

1. Estimating the Sum of 10 Numbers. Students were shown a card with the problem $5 + 5 + 4 + 6 + 3 + 3 + 6 + 6 + 5 + 4 = \square$ and the fact that there were 10 numbers was emphasized. They were then asked, in rapid succession, to estimate the largest and smallest the answer could be, and to give their best guess for the answer.

2. -1 Pattern. Students were shown a card with the problem: $6-5 + 9-8 + 2-1 + 5-4 = \square$ They were then asked to quickly write in the box their guess for the right answer.

3. Reversing Digits. Students were shown the number 1,573 written on a piece of paper and asked to write the nubmer obtained by reversing the 3 and 7. Then they were asked whether the new number was larger or smaller than the original and by about how much. Students were then asked the same questions for reversing the 3 and 5 of the original number, and then reversing the 3 and 1, except they were no longer allowed to write down the new number.

4. Grouping into Hundreds. The students were shown the following problem:

$1.22
1.81
1.51
1.53
1.32
1.33
1.39

They were told "Suppose we had to buy these 7 things and we want to be sure to take enough one dollar bills with us, but we don't want to take any more than we'll need". Then they were asked for a quick estimate, which they could then change after studying the problem. They were also asked to justify their answer.

5. $6 \times \square = 138$. The object was to take turns, using a calculator, in multiplying numbers by 6, until the one producing a product of 138 was found. The results were recorded each time and students encouraged to look at previous trials.

6. $\frac{1}{2}$ of 5,394 = $\square$. Students were asked how they could eventually solve this by using the calculator in a similar manner to #5 (i.e., the "÷" button was excluded). This turned out to be too hard for most students who were then asked to make a guess, and then to check the guess or see if they could find out whether the next guess should be larger or smaller.

7. Seven 69's. Students were shown the following "calendar".

SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
69¢	69¢	69¢	69¢	69¢	69¢	69¢

" It shows that everyday this week Bill is going to get 69¢ for doing chores". Students were then asked what would be the fastest way, on the calculator, to figure out how much Bill would earn by the end of the week."